ASSESSING KINEMATIC RELATIONS WITH HIGH SPEECH RATE RESOLUTION DATA

Stephan R. Kuberski and Adamantios I. Gafos

University of Potsdam, Department of Linguistics and Cognitive Sciences
kuberski@uni-potsdam.de

ABSTRACT

We employ a metronome-driven paradigm of repeated syllable production with systematic speech rate control to assess two important kinematic relations: peak velocity is proportional to movement amplitude, with a proportionality modulated by speech rate, and peak velocity is proportional to the average velocity of a movement. The resulting data show substantially higher kinematic variation than previous works devoted to the same relations. Statistical evaluation of the two relations yields very strong evidence for their presence in the data. Whereas slopes of the first relation are in broad agreement with earlier reports, verified here across a much wider range of speech rates, slopes of the second relation show a rate dependency which has not been previously reported. We consider these results in the context of other work linking slope values to the dynamics underlying the speech movements.

Keywords: articulatory phonetics, speech production, kinematic relations, speech rate

1. INTRODUCTION

In understanding the principles that govern human movements (in general, and speech movements in particular) a large body of previous work has sought to uncover the existence of relations among kinematic parameters such as amplitude (A), peak velocity (v), and duration (T) of movements. In speech, two relations in particular have been at the focus of previous studies: first, peak velocity is proportional to movement amplitude ($v \propto A$), with a proportionality modulated by stress, tone, and speech rate [1–4], and second, peak velocity is proportional to the average velocity of a movement ($v \propto A/T$) [1, 5–7].

Yet evidence for these two relations has remained somewhat limited. A prerequisite to a thorough assessment of any relation is high variation in the parameters to be related. Previous works devoted to the study of the two mentioned relations typically elicited speech movements by conventional verbal instruction (e.g., “speak slow” or “speak fast”). As we demonstrate below, the resulting data exhibit limited kinematic variation. For the present assessment of the two relations, we employed a metronome-driven paradigm of repeated syllable production with systematic, high-range, and high-resolution speech rate control to elicit articulatory movements of much higher, as demonstrated here, kinematic variation.

In the resulting data set, we verify the presence of the two kinematic relations with very strong statistical evidence. The extensive variability elicited in our paradigm furthermore allows an examination of particular rate dependencies of the slope values implicated in the two relations. Regarding the first relation ($v \propto A$), our results verify what is known from previous work, that is, a linear increase in the slope values with increasing speech rate. Regarding the second relation ($v \propto A/T$), however, our results indicate for the first time a declining trend in the slope values as a function of increasing rate. These results and specifically the validity of the two kinematic relations under the rather wide range of speech rate conditions implies the existence of general principles that organize and constrain the functioning of the speech motor system. In the final part of this paper, we thus turn to consider our results in the context of other work drawing bridges between the slope dependency in the $v \propto A/T$ relation and the dynamics underlying the speech movements.

2. METHODS

Five native speakers of German (three female, two male) were recruited at the authors’ institution to participate in an experiment of repeated syllable production. The speakers were between 18 and 27 years old, without any present or past speech or hearing problems. Participation in the experiment was paid and all experimental procedures were performed in compliance with the relevant laws, guidelines and conditions expressed by the authors’ institution.

During the experiment, speakers were asked to produce sequences of repeated /ka/ or /ta/ syllables in time with an audible metronome. The metronome served as an extrinsic index of the intended rate of syllable production with increasing rates between con-

secutive blocks of four trials. Within a trial, speakers were exposed to the beat of the metronome and begun articulation at a point of their choice. Trial lengths (i.e., the number of metronome ticks in a trial) were programmed to allow for the production of 30 consecutive syllables. The rate of the metronome was set to the values of 90, 150, 210, 300, 390, and 480 beats per minute (bpm), covering the range of very slow (90 bpm), slow (down to 150 bpm), normal (around 300 bpm), and fast speech rates (up to 480 bpm); see [8–11] for what ranges of rates are considered normal in German. The experiment begun with the production of /ka/ syllables across all rates, starting with the slowest, and went on with /ta/ syllables afterward.

Data registration was conducted using Electromagnetic Articulography (Carstens AG501). In addition to sensors on rigid reference locations required for head movement correction (left and right mastoids, nose bridge), we tracked the movement of a sensor placed on the tongue body and another on the tongue tip. To preserve the temporal and spatial resolution of the registered data (0.3 mm RMS at 1250 Hz), we employed a for speech not often (e.g., in [12–14]) but in general human movement science frequently used spline smoothing approach. By means of heptic-order splines with a fixed predicted mean square error criterion [15, 16], we obtained signals with more faithful position and velocity characteristics than by traditional filtering and differentiation techniques (see [17], for a comparison of a spline with a filtering approach in speech).

The resulting tongue body (for /ka/) and tongue tip (for /ta/) trajectories were first segmented into movement cycles by means of local tangential velocity minima in the full three dimensions. These minima then served to determine the principal axis of segments defined by their endpoint-to-endpoint vectors in 3-D (referred to as constriction degree in [18] or reaching axis in earlier work [19, 20]). In projecting the three-dimensional movement data onto the principal axes, we finally obtained one-dimensional representations of movements as typically engaged in studies of speech kinematics [6, 21–23]. In a last step, we applied a 20% peak velocity threshold criterion to the data to avoid potential problems with poorly defined transitions into or away from quasi-steady states [24].

In what follows, we refer to movements of positive velocity (motion towards a constriction) as closing movements and to movements of negative velocity (motion away from a constriction) as opening movements; see Figure 1, showing examples of closing and opening movements involved in the production of /ka/ syllables. Each such movement is a period of an effector's continuous motion from

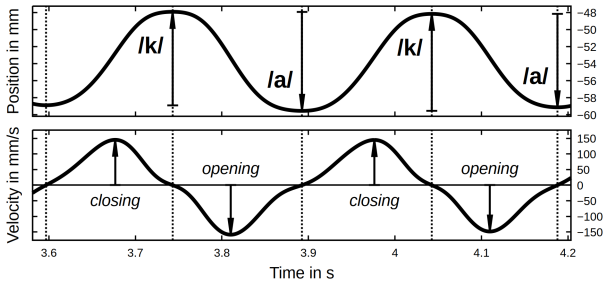

Figure 1: Kinematic trajectories of two /ka/ syllables produced by one of our speakers at normal speech rate. Arrows indicate the kinematic parameters of movement amplitude (top) and peak velocity (bottom).

one quasi-steady state to another, during which the effector's velocity rises from a minimal value up to a maximal value and declines again. Overall, we can identify three temporal landmarks in this period: movement onset, peak velocity, and movement offset. By means of these landmarks, we defined the following kinematic parameters: movement duration (T) as the temporal distance between onset and offset, movement amplitude (A) as the spatial distance between onset and offset, and peak velocity (v) as the value with which velocity peaks within a movement (A and v expressed by the arrows in Figure 1). Across all speakers and metronome rates, we acquired kinematic parameters of 2,470 closing and 2,595 opening movements in the case of /ka/, and 1,948 closing and 2,113 opening movements in the case of /ta/ syllables.

3. RESULTS

We first evaluate the amount of variation elicited in our paradigm by inspecting the range of values in the kinematic parameters defined above. We begin with movement durations, which range from around 45 to 350 ms for /ka/ and from around 35 to 390 ms for /ta/ syllables. Both ranges cover almost one order of magnitude (OM, a tenfold increase in scale), across all speakers and metronome rates.

Figure 2 visualizes movement amplitude and peak velocity as a function of speech rate. Note that, instead of relying on the intended speech rate cued by the metronome, we quantified the actual rate produced by taking the reciprocal value of movement duration (i.e., the abscissae in the figure are in terms of movements per seconds, mps). Data from tongue body movements in /ka/ syllables are shown in the top two panels and data from tongue tip movements in /ta/ syllables are shown in the bottom two panels. As a first result, it becomes clear that, overall, our paradigm successfully elicited movements of a widespread kinematic range: for both syllables,

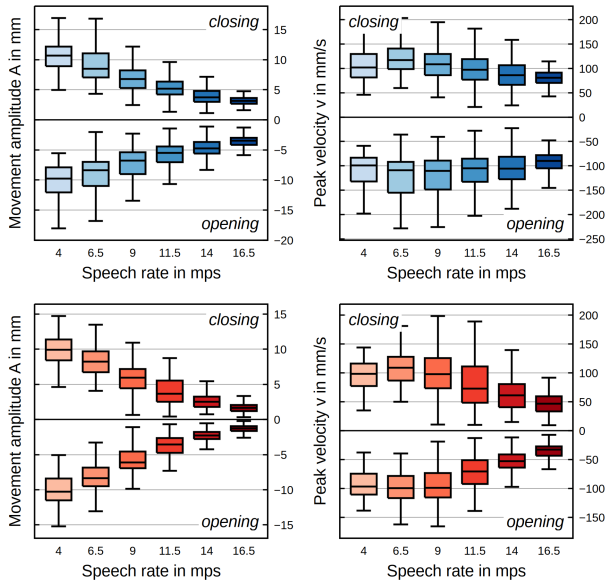

Figure 2: Box plots showing ranges of amplitude (left) and peak velocity values (right) as a function of actual speech rate. Top row shows data of /ka/, bottom row data of /ta/ syllables.

/ka/ and /ta/, the kinematic parameters spread well over the range of an order of magnitude (OM). In particular, for /ka/ syllables, movement amplitudes are in the range of 0.89 to 21 mm (corresponding to 1.4 OM), and peak velocity attains values in the range of 21 to 270 mm/s (1.1 OM); with all ranges given in absolute values. For /ta/ syllables, similar ranges are observed: 0.23 to 16 mm amplitude (1.8 OM), and 7.5 to 210 mm/s peak velocity (1.4 OM).

To facilitate comparison with earlier works [1, 3, 6, 7, 21, 22, 25, 26], we estimated parameter ranges from published tables and graphic representations. Table 1 summarizes the order of magnitudes of these estimates for the ranges of movement duration (T), movement amplitude (A), and peak velocity (v). For example, in a study designed to explore jaw kinematics over a maximally possible range of movement rates, Nelson et al. [25] (second row in the table) reported ranges corresponding to about 0.7 OM in duration, 1.1 OM in amplitude, and 1.0 OM in peak velocity values. Other studies indicate duration ranges in OM of about 0.6 (lower lip) [3, 26], 0.5 (vocal folds) [6], 0.4 (tongue body) [1], and 0.2 (tongue body and tip) [7]; amplitude ranges of about 1.3 (lower lip) [3, 26], 1.1 (tongue body) [21], 0.7 (tongue body) [1], 0.5 (vocal folds) [6], 0.4 (tongue body) [7], and 0.3 (tongue body, tongue tip) [6, 7]; and peak velocity ranges with an estimated OM of about 1.1 (lower lip) [3, 26], 0.7 (tongue body) [21], 0.6 (tongue body) [1], and 0.4 (tongue body, tongue tip) [7]. Overall, then, the kinematic variation achieved by our paradigm (given in the two final rows of Table 1) substantially

Study	Effector	T	A	v
Ostry [21]	tongue body	?	1.1	0.7
Nelson [25]	jaw	0.7	1.1	1.0
Munhall [6]	vocal folds	0.5	0.5	?
	tongue body	?	0.3	?
Ostry [1]	tongue body	0.4	0.7	0.6
Vatikiotis [3, 26]	lower lip	0.6	1.3	1.1
Perkell [7]	lower lip	0.1	0.3	0.3
	tongue body	0.2	0.4	0.4
	tongue tip	0.2	0.3	0.4
present work	tongue body	0.9	1.4	1.1
	tongue tip	1.1	1.8	1.4

Table 1: Comparison of the kinematic variation achieved in different studies. Values which could not be determined from a publication are indicated by a question mark (?). All values in orders of magnitude (OM).

surpasses the majority of ranges in earlier works: order of magnitude values are consistently higher by some tenth of 1 OM. Keeping in mind that an increase of about 0.3 OM corresponds to a doubling of a particular range, it becomes clear that our paradigm elicited movements of almost twice the kinematic variation of previous works in all the kinematic parameters.

We now turn to verify the presence of the two kinematic relations in our data. Figure 3 shows scatter plots of the mutual dependencies between the kinematic parameters as expressed by the two relations, $v \propto A$ (peak velocity is proportional to amplitude, left) and $v \propto A/T$ (peak velocity is proportional to average velocity, right). In the figure, speech rate is color-coded, with fainter shades corresponding to movements of slower, and darker shades corresponding to movements of faster rates. Data from movements of different directions separate into different quadrants of the shown coordinate systems: opening movements reside in the lower left and closing movements in the upper right quadrant (due to our sign convention of the kinematic parameters introduced earlier). Inspection of Figure 3 reveals a clear systematicity, as expected from the two kinematic relations. First, there is a proportionality between peak velocity v and movement amplitude A as expressed by the relation $v \propto A$, with steeper slopes for faster rates (color-coded by different shades). Second, there is a proportionality between peak velocity v and average velocity A/T as expressed by the relation $v \propto A/T$ with a visually indiscernible dependency on speech rate.

To statistically evaluate the empirical evidence for the two kinematic relations, a simple linear regression model without an intercept term was fit to the data. Resulting correlation strengths for /ka/ and /ta/ sylla-

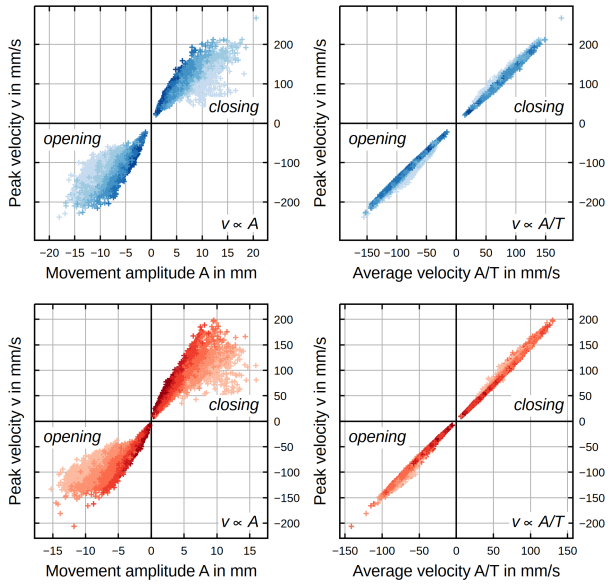

Figure 3: Scatter plots showing the two kinematic relations $v \propto A/T$ (left) and $v \propto A$ (right). Top row shows data of /ka/ syllables, bottom row data of /ta/.

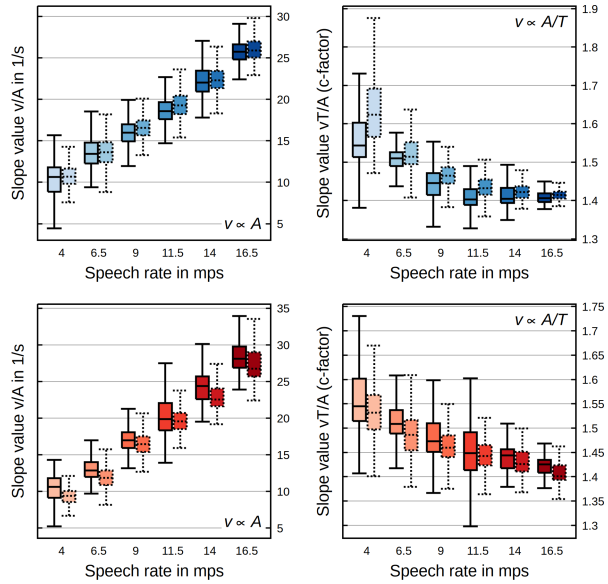

Figure 4: Box plots showing the slope values of the two kinematic relations, $v \propto A$ (left) and $v \propto A/T$ (right). Top row shows data of /ka/ syllables, bottom row data of /ta/. Data of closings (solid) and openings (dotted) are shown side-by-side within each speech rate bin.

bles are in the order of 0.65–0.95 R-squared for $v \propto A$, strictly increasing with speech rate, and 0.95–0.99 for $v \propto A/T$. All p-values reside below 0.001, implying very strong statistical evidence for the presence of the two kinematic relations. Regarding the slopes of the two relations, we evaluated the ratios v/A of each movement for the first, and vT/A for the second relation. Figure 4 then shows the average values as a function of speech rate. For both syllables, these are linearly increasing slopes for the relation $v \propto A$ (left) and declining slopes for the relation $v \propto A/T$ (right). Both speech rate dependencies are significant, with Spearman's rank correlation strengths of 0.99 in case of the first and −0.65 in case of the second relation.

The observed linear increase in the slopes of the $v \propto A$ relation as a function of speech rate is in broad agreement with earlier assessments [1–4], but verified here across a much wider range of speech rates and much higher resolution. Reports on the other $v \propto A/T$ relation [1, 5, 6, 27, 28], however, do not provide any indications of its slope changing with varying speech rate. The literature rather suggests that slope values of the relation remain invariant over a variety of linguistic manipulations including speech rate [1, 5, 6]. However, it must be kept in mind that these studies did not manipulate speech rate to the same extent as in our paradigm.

4. CONCLUSION

In sum, we have employed a paradigm of repeated syllable production with systematic speech rate control to assess two kinematic relations, $v \propto A$ and $v \propto A/T$. The resulting data set shows substantially higher kinematic variation than previous works devoted to the same relations. Statistical evaluation of the two relations yields very strong evidence for their presence in the data. Whereas slopes of the $v \propto A$ relation have been shown to be in broad agreement with earlier reports, slope values of the $v \propto A/T$ relation suggest a rate dependency which has not been previously reported.

The rate dependency of the $v \propto A/T$ relation can be considered in the context of other work which aims to link values of its slope to the dynamics underlying the movements. Specifically, [27] and [5] have pointed out that the relation's slope, referred to as the c-factor, characterizes the shape of a movement's velocity profile (e.g., sinusoidal, trapezoidal, or triangular). In turn, velocity profile shape has served as a diagnostic of the dynamical organization underlying movement. For example, sinusoidal velocity profile shapes are indicative of dynamics corresponding to linear second-order systems, whereas trapezoidal shapes are indicative of dynamics constrained by a minimal impulse of movement [27]. Hence, in future work, we plan to harness the variation in slope values, as uncovered at high resolution in our paradigm, and relate it to the underlying dynamics governing the movements. We further aim to include the less frequently used parameter of peak acceleration which may play out in additional kinematic relations whose slopes, in turn, may provide further information.

ACKNOWLEDGMENT

This work was funded by the German Research Foundation (DFG) under Project ID 317633480, SFB 1287, Project C04. Data collection has been partially funded by the European Research Council (ERC) under Grant agreement ID 249440.

5. REFERENCES

[1] D. J. Ostry and K. G. Munhall, "Control of rate and duration of speech movements," *The Journal of the Acoustical Society of America*, vol. 77, no. 2, 1985.

[2] J. A. S. Kelso, E. Vatikiotis-Bateson, E. L. Saltzman, and B. Kay, "A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling," *The Journal of the Acoustical Society of America*, vol. 77, no. 1, 1985.

[3] E. Vatikiotis-Bateson and J. A. S. Kelso, "Rhythm type and articulatory dynamics in English, French and Japanese," *Journal of Phonetics*, vol. 21, 1993.

[4] B. Kühnert and P. Hoole, "Speaker-specific kinematic properties of alveolar reductions in English and German," *Clinical Linguistics & Phonetics*, vol. 18, no. 6–8, 2004.

[5] K. G. Munhall, "Temporal adjustment in speech motor control: evidence for laryngeal kinematics," PhD thesis, McGill University, Canada, 1984.

[6] K. G. Munhall, D. J. Ostry, and A. Parush, "Characteristics of velocity profiles of speech movements," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 11, no. 4, 1985.

[7] J. S. Perkell, M. Zandipour, M. L. Matthies, and H. Lane, "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues," *The Journal of the Acoustical Society of America*, vol. 112, no. 4, 2002.

[8] A. Gerstenberg, S. Fuchs, J. M. Kairet, J. Schröder, and C. Frankenberg, "A cross-linguistic, longitudinal case study of pauses and interpausal units in spontaneous speech corpora of older speakers of German and French," in *9th International Conference on Speech Prosody*, 2018.

[9] F. Pellegrino, C. Coupé, and E. Marsico, "A cross-language perspective on speech information rate," *Language*, vol. 87, no. 3, 2011.

[10] F. Pellegrino, J. Farinas, and J.-L. Rouas, "Automatic estimation of speaking rate in multilingual spontaneous speech," in *International Conference on Speech Prosody*, 2004.

[11] V. Dellwo and P. Wagner, "Relations between language rhythm and speech rate," in *15th International Congress of Phonetic Sciences*, 2003.

[12] B. J. Kröger, G. Schröder, and C. Opgen-Rhein, "A gesture-based dynamic model describing articulatory movement data," *The Journal of the Acoustical Society of America*, vol. 98, no. 4, 1995.

[13] J. O. Ramsay, K. G. Munhall, V. L. Gracco, and D. J. Ostry, "Functional data analyses of lip motion," *The Journal of the Acoustical Society of America*, vol. 99, no. 6, 1996.

[14] S. Schötz, J. Frid, A. Löfqvist, and L. Gustafsson, "Functional data analysis of tongue articulation in Gothenburg and Malmöhus Swedish /i:, y:, u-:/," in *Fonetik*. Linköping University, 2013.

[15] H. J. Woltring, "A Fortran package for generalized, cross-validatory spline smoothing and differentiation," *Advances in Engineering Software*, vol. 8, no. 2, 1986.

[16] S. R. Kuberski, "GCV spline smoothing in Matlab," 2023. [Online]. Available: github.com/srkuberski/gcvspl

[17] S. R. Kuberski and A. I. Gafos, "Higher-order kinematic analysis of speech movement data," in *14th International Conference on Bio-inspired Systems and Signal Processing*, 2021, conference poster.

[18] E. L. Saltzman and K. G. Munhall, "A dynamical approach to gestural patterning in speech production," *Ecological Psychology*, vol. 1, no. 4, 1989.

[19] E. L. Saltzman and J. A. S. Kelso, "Skilled actions: a task-dynamic approach," *Haskins Laboratories Status Report on Speech Research*, vol. SR-76, 1983.

[20] ——, "Skilled actions: a task-dynamic approach." *Psychological Review*, vol. 94, no. 1, 1987.

[21] D. J. Ostry, E. Keller, and A. Parush, "Similarities in the control of the speech articulators and the limbs: kinematics of tongue dorsum movement in speech," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 9, no. 4, 1983.

[22] D. J. Ostry, J. D. Cooke, and K. G. Munhall, "Velocity curves of human arm and speech movements," *Experimental Brain Research*, vol. 68, no. 1, 1987.

[23] S. G. Adams, G. Weismer, and R. D. Kent, "Speaking rate and speech movement velocity profiles," *Journal of Speech, Language, and Hearing Research*, vol. 36, no. 1, 1993.

[24] P. Hoole, C. Mooshammer, and H. G. Tillmann, "Kinematic analysis of vowel production in German," in *Third International Conference on Spoken Language Processing*, 1994.

[25] W. L. Nelson, J. S. Perkell, and J. R. Westbury, "Mandible movements during increasingly rapid articulations of single syllables: preliminary observations," *The Journal of the Acoustical Society of America*, vol. 75, no. 3, 1984.

[26] E. Vatikiotis-Bateson and J. A. S. Kelso, "Linguistic structure and articulatory dynamics: a cross language study," *Haskins Laboratories Status Report in Speech Research*, vol. SR-103/104, 1990.

[27] W. L. Nelson, "Physical principles for economies of skilled movements," *Biological Cybernetics*, vol. 46, no. 2, 1983.

[28] S. Fuchs, P. Perrier, and M. Hartinger, "A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model," *Journal of Speech, Language, and Hearing Research*, vol. 54, no. 4, 2011.